Nanopublications for Biodiversity Go Live

Lyubomir Penev[‡,§], Tobias Kuhn[|], Richard L. L. Pyle[¶], Daniel Mietchen[#], Seyhan Demirov[¤],
Iva Boyadzhieva[¤], Teodor Georgiev[¤]

‡ Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria
§ Institute of Biodiversity & Ecosystem Research - Bulgarian Academy of Sciences and Pensoft Publishers, Sofia, Bulgaria
| Vrije Universiteit Amsterdam, Amsterdam, Netherlands
¶ Bishop Museum, Honolulu, Honolulu, United States of America
# Ronin Institute of Independent Scholarship, Montclair, United States of America
¤ Pensoft Publishers, Sofia, Bulgaria

Abstract

A nanopublication is the smallest unit of publishable information: a scientifically meaningful
assertion about anything that can be uniquely identified and attributed to its author(s) and
serve to communicate a single statement, its original source (*provenance*) and citation
record (*publication info*) (Mons and Velterop 2009, Kuhn et al. 2021). Nanopublications are
fully expressed in both human-readable and machine-interpretable formats and can be
cited thorugh their unique URI identifiers (IDs).

The development of the concept, formats and use cases of biodiversity-specific
nanopublications started back in 2016 (Penev et al. 2016). Recently, Pensoft and Knowled
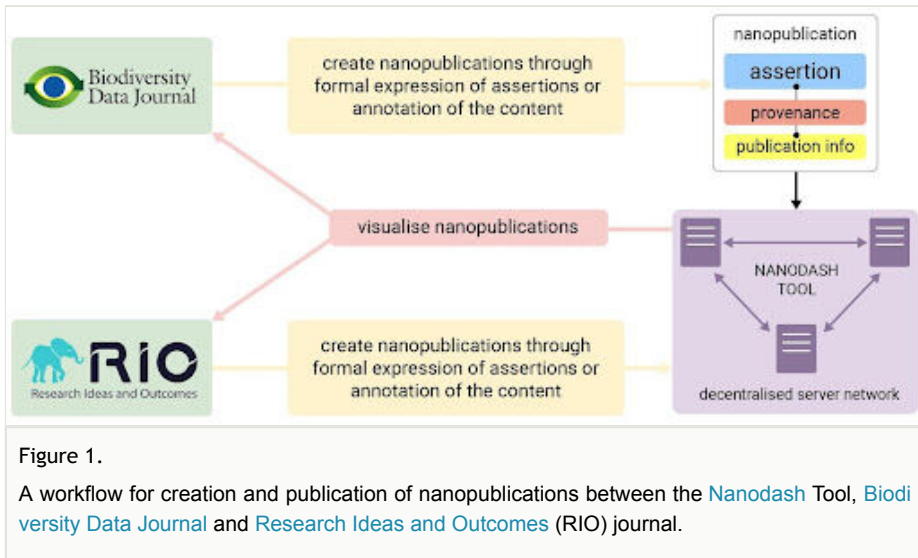ge Pixels created a workflow to publish nanopublications related to a published article
(Fig. 1).

Figure 1.

A workflow for creation and publication of nanopublications between the Nanodash Tool, Biodiversity Data Journal and Research Ideas and Outcomes (RIO) journal.

At its core, the generic biodiversity nanopublication consists of a triple structure with a Subject element, a Relation element, and an Object element. The Subject element refers to a specific organism identified by an Organism ID, or groups of organisms (Taxa, identified by Taxon Concept IDs). Universal Unique Identifiers (UUIDs) are minted for both in all cases, although if the given organism already has an identifier, e.g., from digitised collections or observations, then this identifier can be specified too. The Organism is linked via class membership (rdf:type) to a Taxon Concept, which is identified with another UUID that is minted upon nanopublication creation. The need to mint IDs is due to the fact that Organism or Taxon Concept IDs (which refer to taxon names but also details about their interpretation and use) are not currently available in any comprehensive vocabulary or database. The minted Taxon Concept IDs are further defined by linking them to a Taxon Name ID from Catalogue of Life ChecklistBank (via a relation called "has Taxon Name") and by linking to Taxon Name's concept (interpretation of the name) by a publication's Digital Object Identifier (DOI) or TreatmentBank entry (via a relation called "used as defined in"). Apart from assigning a Taxon Concept, the subject organism can optionally be linked to a life cycle stage (based on the Uber-anatomy (UBERON) Ontology) and to occurrence data via Occurrence ID. The Relation element expresses the association between Subject and Object, according to a particular ontology (e.g., Relation Ontology). The Object element may have different formats. In the template "Association between organisms" the Object has the same structure as the Subject; while in the template "Association between an organism and environment", the Object is a type of habitat available from the Environment Ontolog (ENVO).

In addition to an *assertion*, a nanopublication has a *provenance* record (e.g., DOI of the article where the assertion has been published) and the type of evidence entered in the BasisOfRecord field. The *publication info* always contains the ORCID (Open Researcher and Contributor ID) of the nanopublication's creator, and the timestamp of creation.

Another type of nanopublication template annotates an entire article or selected text from it by using the ORCID of the creator as the Subject, the article's DOI as the Object, and the Citation Typing Ontology (CiTO) to express evaluation of the article (e.g., "supports", "agrees with", "discusses"). These templates are complemented by free text comments and are available under the Nanopub tab on the article webpages at several Pensoft journals (see example and blog).

Nanopublications can also prompt other types of publications, e.g., hypotheses, to become more machine-interpretable, as discussed for invasion biology (Heger et al. 2023).

Keywords

semantic publishing, biodiversity informatics, taxonomy

Presenting author

Lyubomir Penev

Presented at

TDWG 2023

Funding program

Grant title

BiCIKL - Biodiversity Community Integrated Knowledge Library

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Heger T, Jeschke J, Bernard-Verdier M, Musseau C, Mietchen D (2023) The Enemy Release Hypothesis. ARPHA Preprints https://doi.org/10.3897/arphapreprints.e107394

- Kuhn T, Taelman R, Emonet V, Antonatos H, Soiland-Reyes S, Dumontier M (2021) Semantic micro-contributions with decentralized nanopublication services. PeerJ Computer Science 7 https://doi.org/10.7717/peerj-cs.387
- Mons B, Velterop J (2009) Nano-Publication in the e-science era. URL: https://www.w3.org/wiki/images/4/4a/HCLS$$ISWC2009$$Workshop$Mons.pdf
- Penev L, Tuama E, Senverov V, Stoev P, Georgiev T (2016) Nanopublications for biodiversity: concept, formats and implementation. Presentation at TDWG 2016 Conference, Santa Clara, Costa Rica, 4-9 Dec 2016. URL: https://vimeo.com/196875207